DOI : 10.36347/sjebm.2015.v02i02.001

Scholars Journal of Economics, Business and Management
e-ISSN 2348-5302
p-ISSN 2348-8875

Chung-ki Min; Sch J Econ Bus Manag, 2015; 2(2A):139-144

(An International Publisher for Academic and Scientific Resources)

More Accurately Estimating Panel Data Models: A Comparison of Quasi- vs. First-differencing Approaches

Chung-ki Min

Professor of Economics, Dept. of Economics, Hankuk University of Foreign Studies, Seoul, Korea

*Corresponding Author
Chung-ki Min
Email: cmin@hufs.ac.kr or minchungki@gmail.com

Abstract: This study examines whether the consistency and the efficiency can be improved in estimating panel data models when the sample size is realistically finite. In various cases we compare the quasi- and the first-differencing approach which are designed to control for unit-specific effects. The equations transformed by the two approaches are estimated by instrumental variables. Empirical results from simulated data support that the quasi-differencing approach dominates the first-differencing one in estimating and testing panel data models, particularly for small-sized samples.

Keywords: panel data; unit-specific effects; quasi-differencing; first-differencing; orthogonaligy condition; instrumental variable.

INTRODUCTION

Panel data models are widely employed in empirical research because they can control for time- and unit-specific effects. In many panel data the number of cross-sectional units is large and the time period is short. So, a main issue is how to control for unit-specific effects as time-specific effects are easily accounted for by a small number of dummy variables. Two approaches are compared in this study, the first-differencing (FD) and the quasi-differencing (QD) approach. The FD approach, most widely employed one in practice, eliminates the unit-specific effects by subtracting the equation for time period *t*-1 from the one for *t*. However, this approach is valid only when the unit-specific effects are constant over time. The QD approach includes a product term of unit-specific effects multiplied by a time-varying coefficient and then eliminates the product term by a transformation[1-2].

The main objective of this study is to compare the QD and the FD approach with a focus on estimation efficiency. When eliminating the unit-specific effects, we are concerned about a loss of efficiency which could inflate standard errors. The efficiency and thus standard errors are determined by variations in panel data; the unexplained variation in the dependent variable and the variation in the regressors.

In the next section, we present a panel data model and compare the QD and the FD approach along with their instrumental variables. In section 3, after describing the simulated data for various cases, we report and discuss the estimation results. Concluding remarks are provided in section 4.

QUASI- AND FIRST-DIFFERENCING APPROACHES

The model considered in this study is from a two-variable vector autoregressive regression of lag order one, VAR(1). Since the main issue is how to consistently and efficiently estimate a panel data regression model, we focus on only one equation of VAR(1). For cross-sectional unit *i* (=1,..., *M*) and time period *t* (=2, ..., *T*), this model allows for time-specific and unit-specific effects.[1]

$$y_{it} = \alpha_1 y_{i,t-1} + \beta_1 x_{i,t-1} + \delta_t + \psi_t f_i + u_{it} \quad (1)$$

where the error term u_{it} satisfies the orthogonality conditions $E[y_{is}u_{it}] = E[x_{is}u_{it}] = 0 \;\; (s < t)$ and the time-specific effects (δ_t) are common to all cross-sectional units. This model allows for unit-specific effects to vary over time as the

[1] This study considers a dynamic panel data model, but the same discussion also applies to static panel data models.

DOI : 10.36347/sjebm.2015.v02i02.001

time-invariant unit-specific effects f_i are multiplied by a time-varying coefficient ψ_t [2]. In this study we consider two cases of ψ_t; one is time-varying and the other is time-constant ($\psi_t = 1$)[2].

To eliminate the time-varying unit-specific effects, we let $r_t = \psi_t / \psi_{t-1}$ and apply the quasi-differencing transformation suggested by Chamberlain [1]; after multiplying Eq.(1) for time period *t*-1 by r_t, we subtract the result from the equation for time period *t*.

$$y_{it} = \theta_{1t} y_{i,t-1} + \beta_1 x_{i,t-1} + \theta_{2t} y_{i,t-2} + \theta_{3t} x_{i,t-2} + d_t + v_{it} \tag{2}$$

where $\theta_{1t} = \alpha_1 + r_t$, $\theta_{2t} = -\alpha_1 r_t$, $\theta_{3t} = -\beta_1 r_t$, $d_t = \delta_t - r_t \delta_{t-1}$ and $v_{it} = u_{it} - r_t u_{i,t-1}$. The orthogonality conditions of Eq.(1) imply that the error term v_{it} satisfies $E[y_{is} v_{it}] = E[x_{is} v_{it}] = 0$ for $s < t-1$ because of the presence of $u_{i,t-1}$ in v_{it}. Thus, the instrumental variables which can be used to identify the parameters of Eq.(2) are included in the following $1 \times (2t-3)$ vector.

$$z_{it} = [y_{i,t-2}, \cdots, y_{i1}, x_{i,t-2}, \cdots, x_{i1}, 1]$$

Since there are five variables on the right-hand side in Eq.(2), including a constant for d_t, the necessary condition for identification requires that there be at least five instrumental variables, $(2t-3) \geq 5$ i.e., $t \geq 4$ [2]. Thus, the first observation for Eq.(2) starts at $t = 4$.

Using vector notation we can express Eq.(2) more compactly. For time period *t*,

$$Y_t = W_t B_t + V_t \tag{3}$$

where $Y_t = [y_{1t}, \cdots, y_{Nt}]'$ and $X_t = [x_{1t}, \cdots, x_{Nt}]'$ are $M \times 1$ vectors of observations for time period *t*; $W_t = [Y_{t-1}, X_{t-1}, Y_{t-2}, X_{t-2}, 1_M]$ is a $M \times 5$ matrix with 1_M being a $M \times 1$ vector of ones; $V_t = [v_{1t}, \cdots, v_{Nt}]'$ is a $M \times 1$ vector of the transformed disturbances; and $B_t = [\theta_{1t}, \beta_1, \theta_{2t}, \theta_{3t}, d_t]'$ is a 5×1 vector of coefficients. The instrumental variables for time period *t* are $Z_t = [Y_{t-2}, \cdots, Y_1, X_{t-2}, \cdots, X_1, 1_M]$.

We can stack all observations from the entire period.[3]

$$Y = WB + V \tag{4}$$

where $Y = \begin{bmatrix} Y_4 \\ Y_5 \\ \vdots \\ Y_T \end{bmatrix}$, $W = \begin{bmatrix} W_4 & 0 & \cdots & 0 \\ 0 & W_5 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_T \end{bmatrix}$, $B = \begin{bmatrix} B_4 \\ B_5 \\ \vdots \\ B_T \end{bmatrix}$, and $V = \begin{bmatrix} V_4 \\ V_5 \\ \vdots \\ V_T \end{bmatrix}$.

Because of the time-varying coefficients, the orthogonality conditions are defined separately for each *t*.

[2] This coefficient can be set to any constant, only a matter of scale. With $\psi_t = 1$, Eq.(1) becomes a typical panel data model.

[3] The first observation for Eq.(2) starts at *t*=4 because of the necessary condition for identification.

DOI : 10.36347/sjebm.2015.v02i02.001

$$\begin{bmatrix} Z_4^{'}V_4 / M \\ \vdots \\ Z_T^{'}V_T / M \end{bmatrix} \xrightarrow{M\to\infty} 0 \tag{5}$$

Thus, the instrumental variables for Eq.(4) are $Z^{diag} = diag[Z_4, \cdots, Z_T]$ where $diag[\ ]$ denotes a block diagonal matrix with the given entries.

By applying the nonlinear generalized method of moment (GMM), we estimate the lag coefficients (α_1, β_1) and the ratios of the time-varying coefficients of the unit-specific effects $(r_t = \psi_t / \psi_{t-1})$.[4] The transformed time-specific effects, $d_t\ \ (= \delta_t - r_t \delta_{t-1})$, are estimated as the coefficients of time dummies; the original time-specific effects (δ_t) are nuisance parameters and not of our interest.

If ψ_t is constant over time, then $r_t = 1$ and Eq.(2) becomes the first-differenced specification.

$$\Delta y_{it} = \alpha_1 \Delta y_{i,t-1} + \beta_1 \Delta x_{i,t-1} + \Delta\delta_t + \Delta u_{it} \tag{6}$$

where Δ denotes the difference between time period t and t-1. This specification includes two time-invariant coefficients (α_1, β_1) and time-specific effects $(\Delta\delta_t)$. Letting $\Delta Y_t = [\Delta y_{1t}, \cdots, \Delta y_{Nt}]'$, $\Delta X_t = [\Delta x_{1t}, \cdots, \Delta x_{Nt}]'$, $\Delta W_t = [\Delta Y_{t-1}, \Delta X_{t-1}]$, $B = [\alpha_1, \beta_1]'$ and $\Delta U_t = [\Delta u_{1t}, \cdots, \Delta u_{Nt}]'$, we compactly express Eq.(6) for each t.

$$\Delta Y_t = \Delta W_t \cdot B + 1_M \cdot \Delta\delta_t + \Delta U_t \tag{7}$$

The first-differenced time-specific effects $\Delta\delta_t$, which are usually nuisance parameters and controlled for by time dummies, are separated from the time-invariant coefficients. Ignoring the nuisance parameters, the instrumental variables for time period t are $Z_t = [\Delta Y_{t-2}, \cdots, \Delta Y_2, \Delta X_{t-2}, \cdots, \Delta X_2]$. Stacking all observations from the entire period, we let

$$\Delta Y = \begin{bmatrix} \Delta Y_4 \\ \Delta Y_5 \\ \vdots \\ \Delta Y_T \end{bmatrix},\ \Delta W = \begin{bmatrix} \Delta W_4 \\ \Delta W_5 \\ \vdots \\ \Delta W_T \end{bmatrix},\ \Gamma = \begin{bmatrix} \Delta\delta_4 \\ \Delta\delta_5 \\ \vdots \\ \Delta\delta_T \end{bmatrix},\ \Delta U = \begin{bmatrix} \Delta U_4 \\ \Delta U_5 \\ \vdots \\ \Delta U_T \end{bmatrix}, \text{ and } Z^{stack} = \begin{bmatrix} Z_4 \\ Z_5 \\ \vdots \\ Z_T \end{bmatrix}.$$

Then, for the entire period, the first-differenced specification Eq.(6) is expressed as

$$\Delta Y = \Delta W \cdot B + diag[1_M] \cdot \Gamma + \Delta U \tag{8}$$

where $diag[1_M]$ represents time dummies. Since the regressors ΔW are a matrix stacked by each period's ΔW_t, the instrumental variables for ΔW are also formed by stacking each period's Z_t. The instrumental variables Z^{stack} satisfy the orthogonality conditions for the entire period, not period by period.

[4] Since any changes in ψ_t correspond to changes in f_i, individual coefficients (ψ_t) are not identified but only their ratios (r_t) are identified [2].

$$\frac{Z^{stack}\,'\Delta U}{M(T-3)} \xrightarrow{M\to\infty} 0 \tag{9}$$

In panel data the number of time period (T) is usually finite but the number of cross-sectional units (M) is sufficiently large. Thus, consistent estimation becomes possible as M increases to infinity.

The first-differenced specification, Eq.(8), has time-invariant coefficients and its orthogonality conditions are thus defined for the entire observations, Eq.(9). In contrast, since the quasi-differenced specification, Eq.(4), includes time-varying parameters, its orthoganality conditions are defined separately for each time period t, as shown in Eq.(5). So, the QD instrumental variables, $Z^{diag} = diag[Z_4, \cdots, Z_T]$, are block-diagonally formed after the entire observations are divided into each time period. By choosing estimates of time-varying parameters separately for each period, Z^{diag} can satisfy the orthogonality conditions more closely than Z^{stack}. Since the deviations from the orthoganality conditions are used for the calculation of standard errors, the quasi-differencing approach is expected to produce smaller standard errors than the first-differencing approach. Even when $\psi_t = 1$ (also $r_t = 1$) and the coefficients (α_1, β_1) are time-invariant, the quasi-differencing can still outperform the first-differencing for small-sized samples because of sampling errors.[5] Using simulated data, we present empirical evidence favoring the quasi-differencing approach in the next section.

EMPIRICAL RESULTS FROM SIMULATED DATA

To compare the estimation results of the QD and the FD approach, we generate data using the following VAR(1) specification.

$$\begin{aligned} y_{it} &= \alpha_1 y_{i,t-1} + \beta_1 x_{i,t-1} + \delta_t + \psi_t f_i + u_{it} \\ x_{it} &= \gamma_1 y_{i,t-1} + \gamma_2 x_{i,t-1} + w_{it} \end{aligned} \tag{10}$$

Since this study focuses on the first equation on y_{it}, we generate $\{x_{it}\}$ series without including time- and unit-specific effects. After we generate data for t = -19 to 12, we discard the first 20 observations (t = -19 to 0) to minimize any effects of starting values. We assign values for the parameters as follows: $\alpha_1 = 0.7$, $\beta_1 = 0.3$, $\gamma_1 = 0.1$, $\gamma_2 = 0.7$, $\sigma_y = 0.2$, $\sigma_x = 1$, and δ_t's are independently drawn from a uniform distribution between -0.5 and 0.5. When we assign values for the unit-specific effects (f_i), we consider that they need to be correlated with the lagged regressors, $y_{i,t-1}$ and $x_{i,t-1}$. To impose the correlations, we generate values for f_i using a linear relation such that $f_i = 0.05 \times i + \eta_i$ where $\eta_i \sim Uniform(-0.5, 0.5)$. Values for the rest of the parameters are assigned appropriately in the following cases.

After the QD and the FD transformation are applied, we estimate the parameters using the GMM. In doing so, we need to account for serial correlation of the transformed disturbances, $v_{it} = u_{it} - r_t u_{i,t-1}$ in QD and $\Delta u_{it} = u_{it} - u_{i,t-1}$ in FD. Since the main issue of this study is to compare the QD and the FD approach in terms of consistency and efficiency, we skip the serial correlations by using observations from every other period; the observations used for estimation are from t = 4, 6, 8, 10, 12.

[5] Of course, when the sample size is large enough for convergence, the estimates in one period will be the same as the ones in another period. Then use of Z^{diag} has no advantage over use of Z^{stack}.

DOI : 10.36347/sjebm.2015.v02i02.001

Table-1: Estimation results when the coefficient for unit-specific effects is constant at $\psi_t = 1$.

		Quasi-differencing		First-differencing		se(FD) - se(QD)
		estimate	se(QD)	estimate	se(FD)	
Case 1: *M* increases						
	[1] *M*=100 α_1	0.640	0.043	0.679	0.060	0.017
	β_1	0.250	0.033	0.294	0.052	0.019
	[2] *M*=200 α_1	0.624	0.033	0.624	0.041	0.008
	β_1	0.254	0.028	0.242	0.038	0.011
	[3] *M*=500 α_1	0.668	0.027	0.669	0.025	-0.002
	β_1	0.274	0.024	0.271	0.022	-0.001
Case 2: *T* increases						
	[4] *T*=4 α_1	0.751	0.115	0.715	0.113	-0.002
	β_1	0.375	0.079	0.336	0.066	-0.013
	[5] *T*=6 α_1	0.576	0.070	0.586	0.080	0.010
	β_1	0.279	0.053	0.278	0.062	0.009
	[6] *T*=12 α_1	0.640	0.043	0.679	0.060	0.017
	β_1	0.250	0.033	0.294	0.052	0.019

1. The true values for α_1 and β_1 are 0.7 and 0.3, respectively.
2. Number of observations used for estimation: Case 1 (*t*= 4, 6, 8, 10, 12), Case 2 (*M*=100; *t*= 4, 6, .., *T*).

In the following cases, we compare the QD and the FD standard errors when the unit-specific effects are constant at f_i over time with $\psi_t = 1$. In such cases, the FD transformation is appropriate for eliminating the unit-specific effects. However, as shown below, the QD could produce smaller standard errors than the FD for small-sized samples.

(1) Case 1: *M* increases

Row [1] in Table 1 shows that when the number of units (*M*) is finite with 100, the standard errors of QD are smaller than the ones of FD; the differences are 0.017 for α_1 and 0.019 for β_1. As *M* increases to 200 and 500, the differences become smaller.

Even when the parameters are constant over time, their estimates in one period could be different from the ones in another period because of sampling errors for finite-sized samples. If so, the QD is expected to produce smaller standard errors than the FD because Z^{diag} of QD can satisfy the orthogonality conditions more closely than Z^{stack} of FD. Row [1] supports this prediction.

However, when *M* is large enough for convergence, the estimates in one period become the same as the ones in another period. If so, use of Z^{diag} has no advantage over use of Z^{stack}. Row [5] shows that with *M*=500, the differences in standard errors decrease to -0.002 and -0.001. The reason why the QD standard errors are slightly bigger is probably because the QD has more parameters $(r_4, r_6, r_8, r_{10}, r_{12})$ to estimate.

(2) Case 2: *T* increases

For a finite number of units (*M*=100), we compare the standard errors when the number of time periods (*T*) changes. Since the first observations in Eqs.(4) and (8) start at *t*=4 and observations from every other period are used for estimation, the observations in each unit come from *t*=4, 6,..., *T*. So, *T*=4 indicates one observation for each unit (i.e., one time period), *T*=6 indicates two observations, and so on; the maximum number of observations in each unit is five with *T*=12.

As explained above, the QD can produce smaller standard errors than the FD because Z^{diag} of QD has more flexibility in satisfying the orthogonality conditions defined for each time period but Z^{stack} of FD has to satisfy the orthogonality conditions defined for the entire period. When there is only one time period (*T*=4), however, the QD has no more flexibility and thus its standard errors would not be smaller than the ones of FD. Row [4] in Table 1 shows that the standard errors of QD are even bigger than the ones of FD; the differences are -0.002 and -0.013. The reason why the QD standard errors are slightly bigger is probably because the QD has one more parameter $r_4(=\psi_4/\psi_3)$ to estimate.

As *T* increases to 6, 8, 10 and 12, the standard errors of both QD and FD become smaller with more observations, but the QD standard errors decrease more sharply than the FD ones; the differences of their standard errors increase from (0.010, 0.009) to (0.017, 0.019). These results support the above explanation that the QD has more flexibility in satisfying the orthogonality conditions because they are defined separately for each time period. Therefore, the advantage of QD over FD becomes bigger as the number of periods increases.

CONCLUSIONS

We have examined whether the consistency and the efficiency can be improved in estimating panel data models when the sample size is realistically finite. In this study we have applied the quasi-differencing (QD) approach by assuming that unit-specific effects vary over time. We have also applied the first-differencing (FD) approach which is valid only when unit-specific effects are constant over time. Overall, the QD approach has produced more accurate estimates than the FD approach in all cases considered in this study.

ACKNOWLEDGMENT

This work was supported by the Hankuk University of Foreign Studies Research Fund of 2014.

REFERENCES

1. Chamberlain G; Panel data, Ch. 22 in The Handbook of Econometrics Volume II, Ed. by Z. Griliches and M. Intrilli gator, Amsterdam: North-Holland Publishing Company. 1983.
2. Holtz-Eakin D, Newey W, Rosen HS; Estimating vector autoregressions with panel data. Econometrica, 1988; 56 (6) :1371-1395.
3. Arellano M, Bond S; Some tests of specification for panel data: Monte Carlo evidence and an application to employ ment equations. Review of Economics Studies, 1991; 58:277-297.
4. Arellano M, Bover O; Another look at the instrumental variables estimation of error components models. Journal of Econometrics, 1995; 68(1):29-52.